ÁRVORES BINÁRIAS ALEATÓRIAS

YOSHIHARU KOHAYAKAWA

1. Árvores binárias

1.1. **Árvores binárias e o número delas.** Começamos com a definição de árvores binárias.

Definição 1. *Uma árvore binária pode ter apenas um nó externo e nenhum nó interno, ou pode consistir de um nó interno r, chamado raiz, e duas subárvores binárias: a subárvore esquerda T_L de r e a subárvore direita T_R de r; essas subárvores são também árvores binárias. A raiz de T_L é o filho esquerdo de r e a raiz de T_R é o filho direito de r. Naturalmente, o nó r é o pai de seus filhos.*

Há uma árvore com 0 nós internos, uma com 1 nó interno, e duas com 2 nós internos. As cinco árvores com 3 nós internos aparecem na Figura 1.

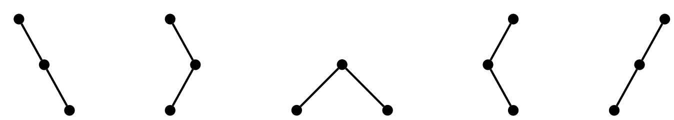

FIGURA 1. As 5 árvores binárias com 3 nós internos (nós externos não estão representados nos diagramas)

Na Tabela 1, temos o número de árvores binárias t_N com N nós internos para valores pequenos de N.

N	1	2	3	4	5	6	7	8	9	10
t_N	1	2	5	14	42	132	429	1430	4862	16796

TABELA 1. Número de árvores binárias t_N com N nós internos

Exercício 2. *Desenhe todas as 14 árvores binárias com 4 nós internos.*

Teorema 3. *O número de árvores binárias t_N com N nós internos é dado por*

$$t_N = \frac{1}{N+1}\binom{2N}{N} = \frac{4^N}{\sqrt{\pi N^3}}\left(1 + O\left(\frac{1}{N}\right)\right). \tag{1}$$

Date: 16 de maio de 2007.

1.2. **Profundidade média de um nó em uma árvore.** Definimos agora a profundidade de um nó em uma árvore.

Definição 4. *A profundidade* $\mathrm{prof}(x)$ *de um nó* x *em uma árvore é* 0 *se* x *é a raiz da árvore e é* $\mathrm{prof}(p)+1$ *se* x *tem pai* p.

Dada uma árvore binária T, estamos interessados na profundidade média de seus nós:

$$\mathrm{Ave}_x(\mathrm{prof}(x)) = \frac{1}{N}\sum_x \mathrm{prof}(x), \tag{2}$$

onde a soma é sobre todo nó interno x de T e N é o número de nós internos em T.

2. Árvores binárias aleatórias

2.1. **Árvores aleatórias uniformes.** Fixe um natural N. Podemos gerar uma árvore binária aleatória com N nós internos simplesmente sorteando, uniformemente ao acaso, uma dentre as t_N árvores binárias com N nós internos. Vamos denotar tal árvore aleatória por U_N.

2.2. **Árvores aleatórias de busca.** Podemos gerar uma árvore binária aleatória com N nós internos da seguinte outra forma.

Podemos usar árvores binárias para armazenar uma coleção de dados, armazenando um dado por nó interno. Neste contexto, supomos que os dados armazenados admitem uma ordem total e que os dados são armazenados de acordo com a *propriedade de árvore binária de busca*: suponha que o dado armazenado em um dado nó x é D; então D é maior que os dados armazenados nos nós da subárvore esquerda de x e D é menor que os dados armazenados nos nós da subárvore direita de x. Quando usamos árvores binárias como tal estrutura de dados, obtemos o que é usualmente conhecido como *árvores binárias de busca*.

Para gerar uma árvore binária aleatória com N nós internos, podemos gerar uma permutação aleatória σ de $\{1,\dots,N\}$ uniformemente ao acaso, e então inserir $\sigma(1),\sigma(2),\dots$ em uma árvore binária de busca inicialmente vazia (um dado inserido gera um novo nó no "fundo" da árvore). Tal processo gera uma árvore binária de busca aleatória com N nós internos e, se ignoramos os dados armazenados em tal árvore, temos uma árvore binária aleatória. Denotemos tal árvore aleatória por B_N.

Naturalmente, o processo descrito acima induz uma distribuição de probabilidade sobre as t_N árvores binárias com N nós internos. Essa distribuição não coincide com a distribuição uniforme (para $N \geq 3$).

Exercício 5. *Qual é a distribuição de probabilidade induzida pelo processo acima sobre as* 5 *árvores da Figura 1? Qual é a distribuição induzida sobre as* $t_4 = 14$ *árvores binárias com* 4 *nós internos?*

3. Profundidade média de nós em árvores aleatórias

Estamos interessados na profundidade média dos nós internos em árvore aleatórias (veja (2)). De fato, se usamos uma árvore binária como uma árvore binária de busca, essa profundidade média é o valor esperado do número de comparações necessárias para localizar um dado armazenado na árvore, supondo que fazemos busca dos dados de acordo com a distribuição uniforme (a probabilidade de buscarmos um particular dado é $1/N$ se há N dados armazenados na árvore).

Lembre que definimos as árvores binárias aleatórias U_N e B_N (veja Seções 2.1 e 2.2).

Teorema 6. *O valor esperado da profundidade média dos nós internos de U_N é*

$$\left(1+\frac{1}{N}\right)4^N\binom{2N}{N}^{-1}-3-\frac{1}{N}=\sqrt{\pi N}-3+O(1/\sqrt{N}). \tag{3}$$

Teorema 7. *O valor esperado da profundidade média dos nós internos de B_N é*

$$2(H_{N+1}-1)\left(1+\frac{1}{N}\right)-2=2\log N-2(2-\gamma)+O\left(\frac{\log N}{N}\right), \tag{4}$$

onde $H_{N+1}=1+1/2+\cdots+1/(N+1)$ é o $(N+1)$-ésimo número harmônico e $\gamma=0.5772156694\ldots$ é a constante de Euler.

Vemos de (3) e (4) que a profundidade média dos nós em U_N é *muito maior* que a profundidade média dos nós em B_N: a primeira é $\sim\sqrt{\pi N}=\Theta(\sqrt{N})$ e a segunda é $\sim 2\log N=\Theta(\log N)$.

Exercício 8. *Seja C_N o valor esperado da profundidade média dos nós internos de B_N. Prove que $C_1=0$ e que, para $N>0$,*

$$NC_N=N-1+\frac{1}{N}\sum_{1\le k\le N}\big((k-1)C_{k-1}+(N-k)C_{N-k}\big). \tag{5}$$

Exercício 9. *Prove que C_N é dado pela equação (4).*

4. Algumas árvores aleatórias

Para visualizarmos a diferença dramática entre (3) e (4), podemos gerar algumas U_N e algumas B_N e podemos desenhá-las. As seis árvores a seguir tem $N=400$ nós internos. Três delas foram geradas com a distribuição de U_{400} e as outras três foram geradas com a distribuição de B_{400}.

Exercício 10. *Escreva um programa que, dado N, gera uma B_N.*

Exercício 11. *Escreva um programa que, dado N, gera uma U_N.*

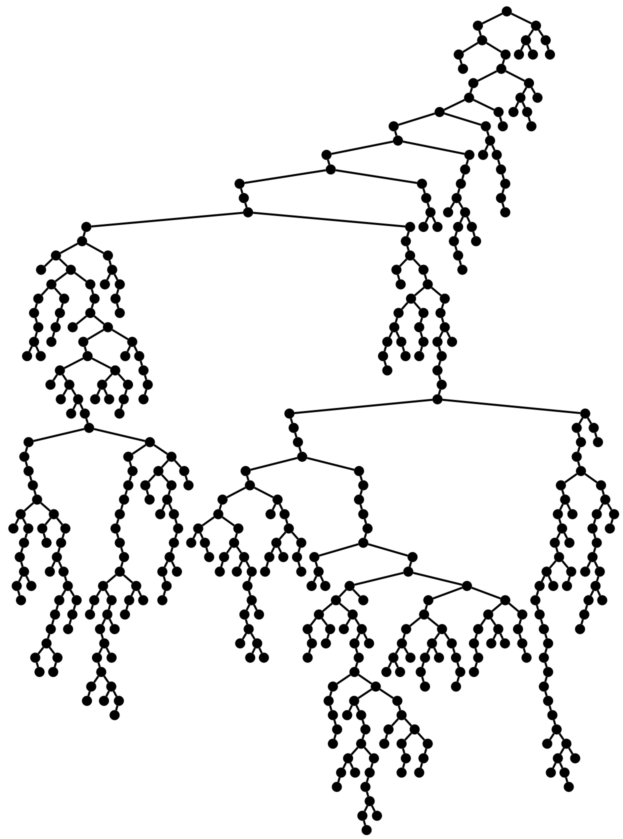

FIGURA 2. Uma U_{400}

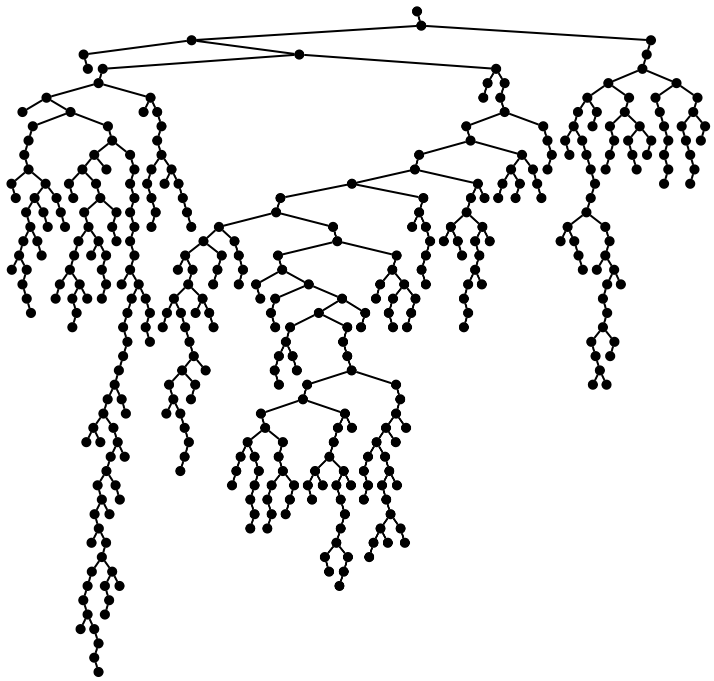

Figura 3. Outra U_{400}

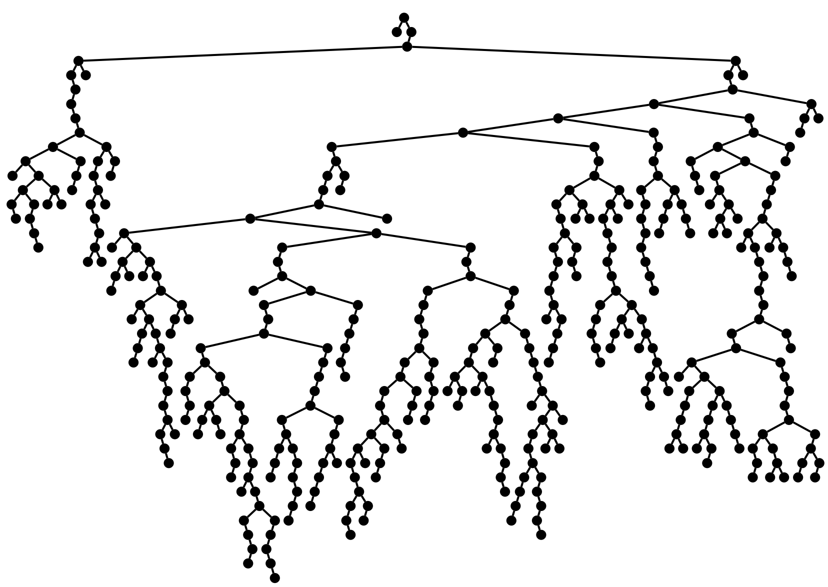

Figura 4. Outra U_{400}

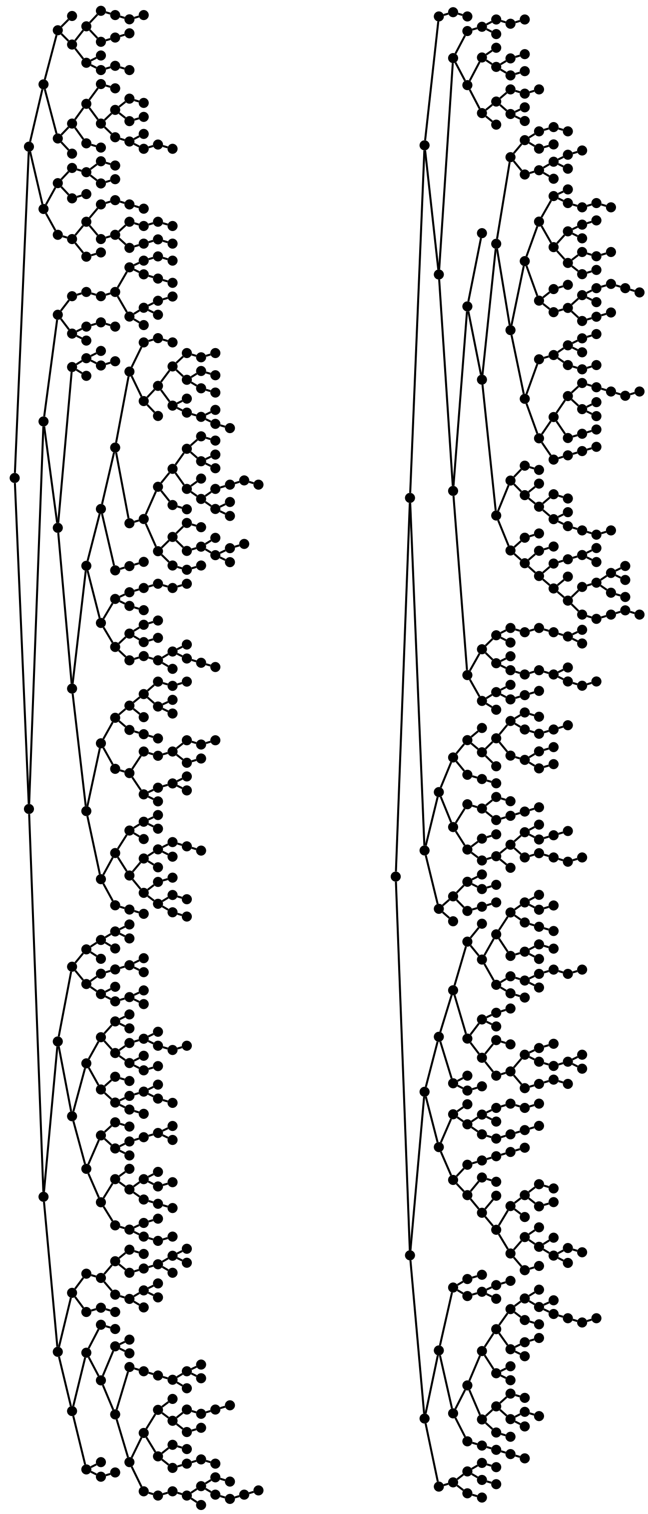

FIGURA 5. Duas B_{400}

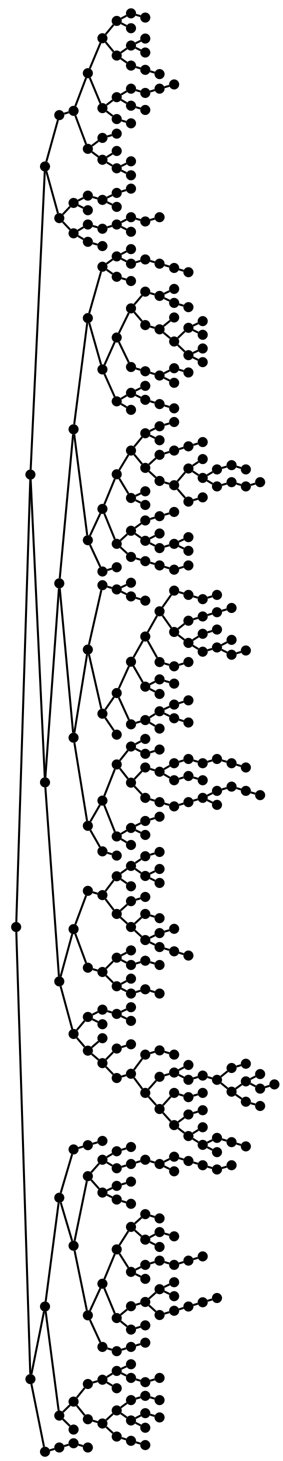

FIGURA 6. Outra B_{400}